Recommendation Algorithms and Evaluation

Group8

April 12, 2016

- Social Network Analysis & Clustering
- Biclustering Analysis
- Chi-Square Analysis
- Model Evaluation

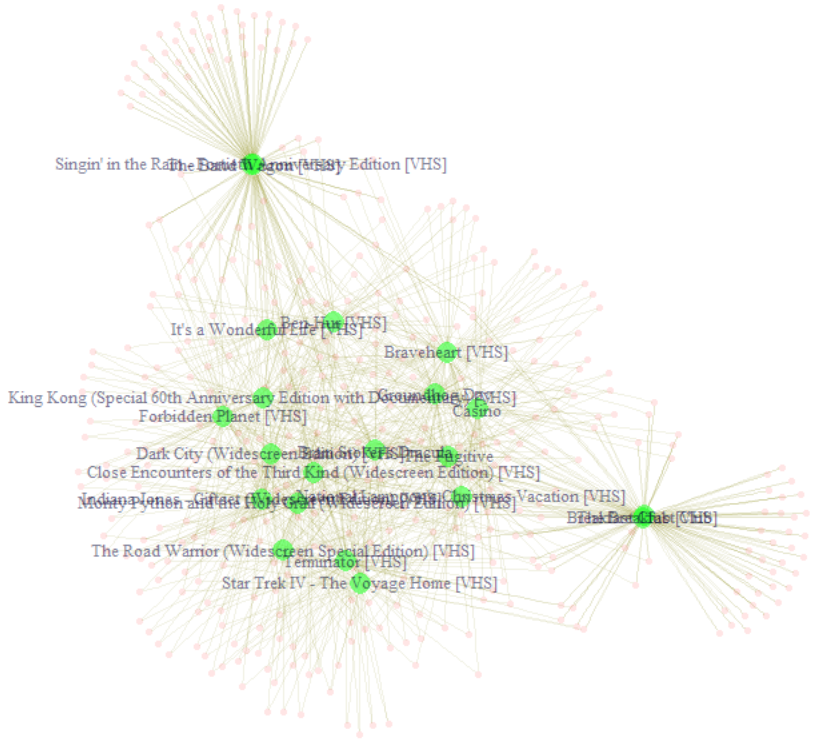
1 Social Network Analysis & Cluster

In order to better understand movie-user biparite network, the following parts fisrt visualize a sample biparite network and show some primary characteristics. Then by converting the biparite network to one-mode movie network, we can have a deeper understanding on movie grouping. Furthermore, based on the idea of movie grouping, we introduced fast greedy algorithm to deal with our large-scale network. And finally we are able to reach our movie recommendation system.

1.1 Biparite Network:

```
whole_data <- readRDS('data_use.RDS')
df1 <- data.frame( user=whole_data[1:2000,]$review_userid, movie=whole_data[1:2000,]$
product_name, stringsAsFactors = F)
m1 <- table( df1 )
user_movie1 <- as.matrix( m1 )

oldw <- getOption("warn")
options(warn = -1)

i2mode<-graph.incidence(user_movie1, mode=c('all'))
V(i2mode)$color[1:1150] <- rgb(1,0,0,.5)
V(i2mode)$color[1151:1171] <- rgb(0,1,0,.5)

V(i2mode)$label <- V(i2mode)$name
V(i2mode)$label.color <- rgb(0,0,.2,.5)
V(i2mode)$label.cex <- .4
V(i2mode)$size <- 6
V(i2mode)$frame.color <- NA

E(i2mode)$color <- rgb(.5,.5,0,0.2)
E(i2mode)$width <- 0.5

# simplied 2 mode graph (degree > 1)
i2mode <- delete.vertices(i2mode, V(i2mode)[degree(i2mode)==1])
```

```
V(i2mode)$label[1:427] <- NA
V(i2mode)$color[1:427] <-  rgb(1,0,0,.1)
V(i2mode)$size[1:427] <- 2

E(i2mode)$width <- .3
E(i2mode)$color <- rgb(.5,.5,0,.1)

options(warn = oldw)

plot(i2mode, layout=layout.fruchterman.reingold)
```



This is a small sample biparite network. Green node means moive. Red node means reviewer. Edge means there is a review from a reviewer to a movie. The position can show the closeness of movies in some sense.

1.2 One-Mode Movie Network

```
##########movie overlap

gmovie_mode1 <- tcrossprod(t(user_movie1))
olmovie <- gmovie_mode1/diag(gmovie_mode1)
olmovie[is.na(olmovie)] <- 0

magdiag <-diag(olmovie)
magallg <- graph.adjacency(olmovie, weighted=T)

# Degree
V(magallg)$degree <- degree(magallg)

# Betweenness centrality
V(magallg)$btwcnt <- betweenness(magallg)



magall<-olmovie

magallgt1 <- magall
magallgt1[magallgt1<0.2] <- 0
magallggt1 <- graph.adjacency(magallgt1, weighted=T)

# Removes loops:
magallggt1 <- simplify(magallggt1, remove.multiple=FALSE, remove.loops=TRUE)
magallggt1$layout <- layout.fruchterman.reingold(magallggt1)
V(magallggt1)$label <- V(magallggt1)$name



# Set vertex attributes
V(magallggt1)$label <- V(magallggt1)$name
V(magallggt1)$label.color <- rgb(0,0,.2,.6)
V(magallggt1)$size <- 6
V(magallggt1)$frame.color <- NA
V(magallggt1)$color <- rgb(0,0,1,.5)

# Set edge attributes
E(magallggt1)$arrow.size <- .3

# Set edge gamma according to edge weight
egam <- (E(magallggt1)$weight+.1)/max(E(magallggt1)$weight+.1)
E(magallggt1)$color <- rgb(.5,.5,0,egam)


plot(magallggt1)
```

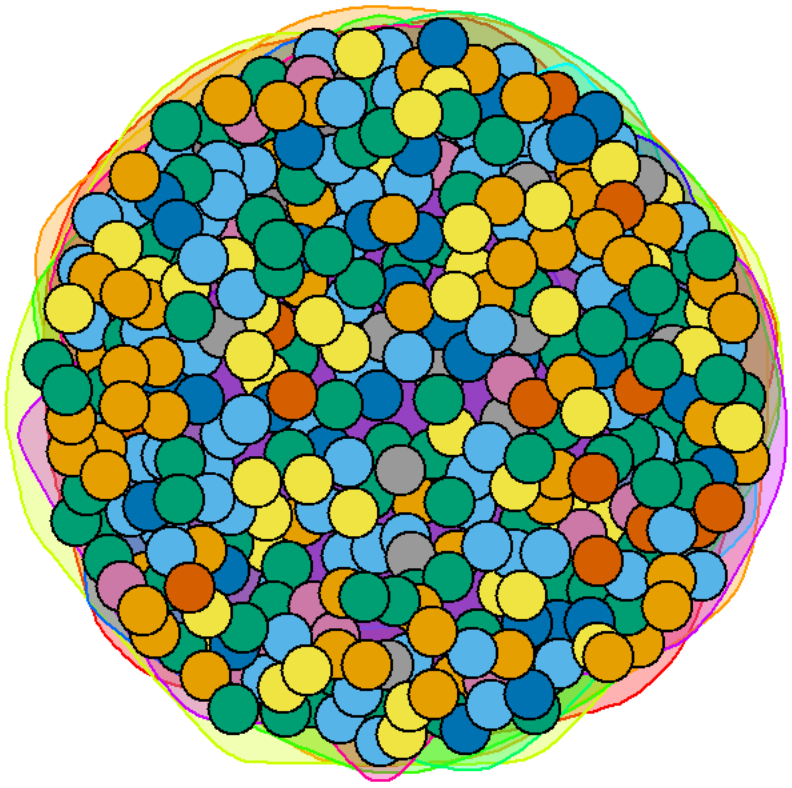
This network is One-Mode Movie Network, which also shows a strong overlap of reivewers between two movies. If there is an arrow from movie A to movie B, at least 20% of reviewers of movie A also have given reviews to movie B. This phenomenon can indicate that we can recommend movie B to the audience of movie A. This method is meaningful but pretty slow. Thus we need to apply other methods with the similar idea.

1.3 Community Dection:

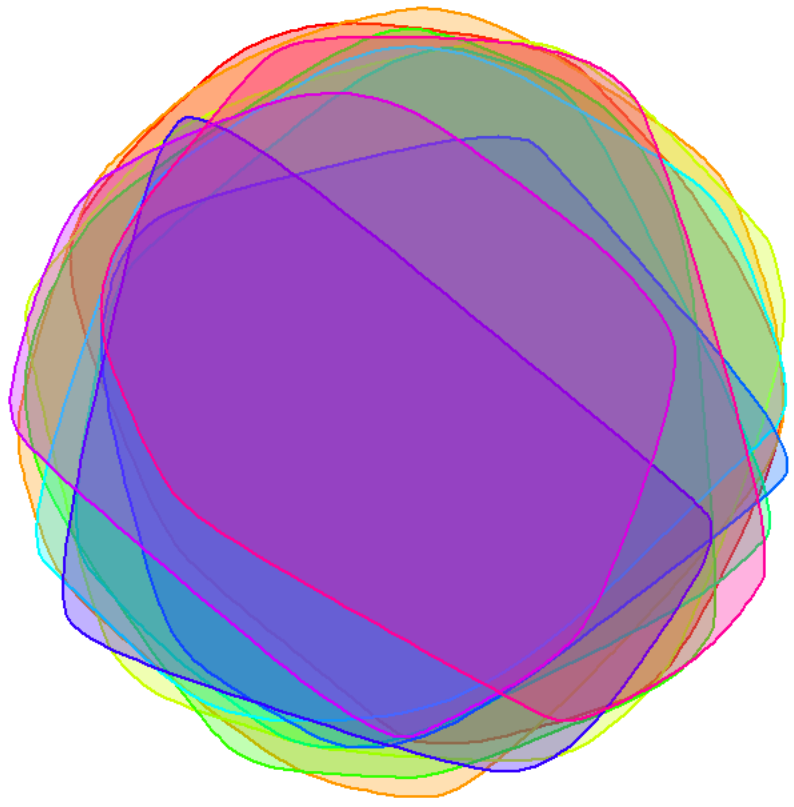
Faster Greedy Algorithm tries to find dense subgraph, also called communities in graphs via directly optimizing a modularity score. Modularity designed to measure the strength of division of a network into modules. Networks with high modularity have dense connections between the nodes within modules but sparse connections between nodes in different modules.

For one-mode movie network, nodes represent movies. Edge means that there is at least one common reviewer for the two movies connected. The weight of an edge means the number of common reviewers for two movies connected. Therefore, a community in network means movies receiving attention from a closed group of people, which shows the closeness of movies.

Faster Greedy Node: Node represents movie; Color of node represents community.



Faster Greedy Group: Color of area means cluster. Below is the size of each cluster



```
## Community sizes
##  1  2  3  4  5  6  7  8  9 10
## 45 67 85 44 32 13 12 11 15 10
```

1.4 Recomendation Algorithm 1:

This recommendation algorithm based on community detection of movie network and ranking of movie score. Intuitively, the result will be a good movie fitting people with similar taste.

Advantage: 1. considered attention of movies from people with similar taste 2. considered quality of movies 3. fast

Disadvantage: 1. didn’t consider past watching experience of the user 2. didn’t include randomness into recommendation

```
# input
movie_name='American Werewolf in London [VHS]'
```

```
for (i in 1:length(fc)) {
  idx=match(movie_name,unlist(fc[i]))
  if(!is.na(idx)){
    break
  }
}
recommend_list=unlist(fc[i])
recommend_list=recommend_list[recommend_list!=movie_name]

score=rep(0,length(recommend_list))

for (i in 1:length(recommend_list)) {
  if(substr(movie_name,1,4)==substr(recommend_list[i],1,4)){
    recommend_list=recommend_list[recommend_list!=recommend_list[i]]
  }
  else{score[i]=mean(subset(movie_user_data,recommend_list[i]==movie_user_data$produc
t_name)$review_score)}

}

op=data.frame(cbind(recommend_list,score))
op<-op[order(op$score,decreasing=TRUE),]
op[1:3,]
```

```
##                                                  recommend_list
## 102            Jaws - 25th Anniversary Collector's Edition [VHS]
## 103 Jaws (25th Anniversary Widescreen Collector's Edition) - DTS
## 108            Jaws (Widescreen Anniversary Collector's Edition)
##                score
## 102 4.86206896551724
## 103 4.83908045977012
## 108 4.82758620689655
```

Output: top3 movies and their corresponding scores